Discussion Paper No. 1077

SUBJECTIVE GAMES AND EQUILIBRIA: I+

by

EHUD KALAI*

and

EHUD LEHRER*

August 1993

Presented at the 1993 Nobel Symposium on Game Theory, Björkborn, Sweden. This paper is an extended version of "Bounded Learning Leads to Correlated Equlibrium" (see Kalai and Lehrer (1991).

---

* Department of Managerial Economics and Decision Sciences, J.L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208.

+ The authors wish to acknowledge valuable conversations with Eddie Dekel-Tabak and other seminar participants of the 1993 Summer in Tel Aviv Workshop, the University of California, San Diego, the California Institute of Technology, and the University of Chicago. This research was supported by NSF Economics, Grant Nos. SES-9022305 and SBR-9223156, and by the Division of Humanities and Social Sciences of the California Institute of Technology.

SUBJECTIVE GAMES AND EQUILIBRIA: I[+]

by

Ehud Kalai[*]

and

Ehud Lehrer*

August 1993

---

[*]Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208.

[+]The authors wish to acknowledge valuable conversations with Eddie Dekel-Tabak and other seminar participants of the 1993 Summer in Tel Aviv Workshop, the University of California, San Diego, the California Institute of Technology, and the University of Chicago. This research was supported by NSF Economics, Grant Nos. SES-9022305 and SBR-9223156, and by the Division of Humanities and Social Sciences of the California Institute of Technology. This is an extended version of the paper entitled "Bounded Learning Leads to Correlated Equilibrium" (see Kalai and Lehrer (1991)).

Abstract

SUBJECTIVE GAMES AND EQUILIBRIA

by Ehud Kalai and Ehud Lehrer

Applying the concepts of Nash, Bayesian or correlated equilibrium to analysis of strategic interaction, requires that players possess objective knowledge of the game and opponents' strategies. Such knowledge is often not available.

The proposed notions of subjective games, and subjective Nash and correlated equilibria, replace unavailable objective knowledge by subjective assessments. When playing such a game repeatedly, subjective optimizers will converge to a subjective equilibrium. We apply this approach to some well known examples including a single multi-arm bandit player, multi-person multi-arm bandit games, and repeated Cournot oligopoly games.

1. Introduction

The notions of a subjective game and subjective equilibria, formulated in this paper, model strategic interaction in uncertain complex dynamic environments. In such situations, players often possess only partial knowledge of the game. Therefore, classical game theoretic approaches, where a significant amount of objective knowledge is assumed, are unrealistic, even for fully rational players. Under the subjective approach, proposed in this paper, each individual player replaces missing objective knowledge by subjective assessments, which he uses in computing an optimal strategy. Following these subjectively optimal strategies, the players eventually converge to a subjective equilibrium.

The proposed subjective model is drastically different from the Nash (1950) and Harsanyi (1967) models. In Nash's formulation, precise, detailed information, like the set of opponents and their strategies, is assumed to be known to every player. In Harsanyi's extension of Nash to Bayesian games, each player assigns the objective correct probability distribution to all conceivable games that may be played, and within each such game he knows the set of opponents and their strategies.

The subjective model departs from Nash and Harsanyi in two important ways. First, players replace missing knowledge by subjective assessments which are not assumed to be correct nor to coincide with each other's. Second, an individual player does not attempt to assess the complete game, i.e., nature's moves, the set of all possible opponents, and their strategies. He restricts himself to aggregate data sufficient for computing his best strategy. In other words, the player views his strategy choice in the game as a one person decision problem.

Nevertheless, the proposed subjective approach can be also thought of as an extension of Nash or Harsanyi. When the subjective assessments of the players are

sufficiently complete and coincide with the true game and chosen strategies, the subjective equilibria proposed coincide with the corresponding (objective) Nash equilibria.

To separate the objective and the subjective data, our model of an n-person *subjective game* consists of two components. The first is a standard n-person infinitely repeated stochastic-outcome game with discounted future payoffs and partial monitoring. It describes the actual game that will be played, i.e. the real strategies and information available to the individual players throughout the game, as well as their payoff functions. However, since the players do not fully know this real game, a second component describes their individual subjective conjectures regarding payoff-relevant events in their own portion of the game.

These conjectures are described by each player through an individual environment response function. Such a function assigns probabilities to all the individual outcomes he may encounter in every stage of the game and for every action he may take. In other words, it is the individual decision tree that he believes to be encountering in the real game by the opponents strategies, whatever they may be. Clearly, the real game and actual opponents' strategies induce on him an objective environment response function and it is unlikely that his subjective function coincides with the objective one.

The real game and the vector of n subjective environment response functions form the subjective game. A vector of n-strategies of the real game is *subjectively rational* if each individual's strategy is optimal relative to the individual's subjective environment response function. The vector is a *subjective equilibrium* if in addition to being subjectively rational, it has a *belief-confirmation* property. That is, the subjective probability assigned by a player to any event observable to him in the play of the game coincides to the true probability induced by the real game and chosen

strategies.

Our convergence result gives a long run justification to the belief-confirmation property. Assuming that the players start with subjectively rational strategies, under sufficient conditions relating their beliefs with the truth, Bayesian updating will lead them eventually to beliefs which are confirmed. In other words, they will converge to a subjective equilibrium.

Before continuing with the general model, and its relationship to earlier literature, we illustrate our approach and concepts through an n-person, infinitely-repeated Cournot game with differentiated products. This game may be thought of as a multi-product extension of the Porter (1983) model, used later by Green and Porter (1984) and by Abreu, Pearce and Staccheti (1986).

At the beginning of every period, each of the n-producers decides on a non-negative quantity of his own good to be produced for the coming period. For each fixed n-vector of chosen production levels, there is a fixed probability distribution determining a random vector of n individual prices for the n-producers. Each producer is informed of his price realization, with which he can compute his period's profit. The game continues in this manner where, prior to every period, each producer's strategy may depend on his own history of past production levels and realized prices. There is a high level of imperfect monitoring here since a player sees only his own realized prices. But even if he saw the prices of others he may still not be able to infer opponents' quantities. The general model presented in the body of the paper allows for a large variety of information systems, ranging from perfect, full information to only learning one's own payoff.

In order to explain and contrast the subjective approach with existing objective ones, we consider the problem of determining an optimal strategy from a producer's viewpoint. The uncertainties faced by him are many. He must know the

market demand function. This demand function depends, however, on the production levels of all his competitors. For this purpose he must identify all related products. He must therefore know the competitors capable of producing these related products, their production capabilities, their information structures, their utility functions, etc. For example, a soft drink producer must have objective knowledge regarding production possibilities, information systems, and strategies of all other soft drink producers and of producers of related products--e.g., fruit juice, milk, and their related products. A naive version of Nash approach will assume that the player and all his opponents know major parameters of this complex game and with this common knowledge somehow contemplate their way to one selected Nash equilibrium.

The above assumption, that all the ingredients of the game are known, seems non-realistic here. In order to improve it one may try to resort to Harsanyi's extension to Bayesian games. In this model the players consider all the possible values of parameters of the above unknown game, and possess common knowledge of the prior probability distribution by which nature selected the actual game played. Within each possible game each player somehow selects a strategy which is best response to the profile of strategies selected by all various types of his opponents in all possible games. Thus by some process of contemplation players arrive to one large vector of strategies, which is a Bayesian equilibrium of the giant game. This concept seems even less realistic since it requires objective knowledge over a much larger space. In addition, it stresses the rationality assumption to an unrealistic level for modeling people's behavior.

The subjective approach to the problem, which we proceed to describe now, also makes non-realistic assumptions on the knowledge and rationality of the players. However, it is less demanding than existing models and, in this sense, presents a

move in the right direction.

Rather than attempting to model the parameters of all potential producers of related products, the subjective player will be assessing only the environment response function, induced on him by the real game and real opponents. Such a function specifies a probability distribution over the prices he may realize, in every stage, and for every one of his production levels. Two facts are important to notice. First, the real game and opponents' strategies induce on him a real environment response function which is usually unknown to him. Second, whatever the real game and opponents' strategies are, finding an optimal strategy in the game is equivalent to finding an optimal strategy relative to the induced environment response function. In other words, the environment response function summarizes all the payoff relevant uncertainties of the game into a one person decision problem.

It follows that assessing the environment response function, instead of the game and equilibrium strategies, involves no loss of generality. In addition, on practical grounds, the environment response function may be easier to assess. It is defined over a drastically smaller space. Moreover, many different games give rise to the same environment response function and they may be considered as one.

The above analysis leads to the incorporation of the subjective environment response functions into the formal model. A subjective version of our Cournot game consists of the real repeated Cournot game, with an n-vector of subjective environment response functions assessed by the individual players.

A vector of strategies in the above game is subjectively rational if each player strategy, i.e., dynamic production plan, is optimal against his subjective environment response function, i.e., his conjectured price responses to quantities produced by him at different stages. In a subjective Nash equilibrium, he has belief-confirmation in addition to the subjective optimization above. This means that, for his chosen

production levels, the subjective probabilities he assigns to realized prices coincide with the objective probabilities, i.e., the ones generated by the competitors chosen strategies in the actual market.

While the belief-confirmation condition just stated is non-realistic for interactions that have just started, it is natural for long, ongoing interactions. Indeed, our convergence result describes sufficient conditions under which subjective optimizers must converge with time to play a subjective equilibrium. Due, however, to the possibility of imperfect monitoring, the limit may be a subjective correlated equilibrium (see Aumann (1974, 1987), Fudenberg and Tirole (1992), and Myerson (1991)) rather than subjective Nash equilibrium. The past play that has lead them to equilibrium turns out to serve as a natural, unavoidable correlation device (see Lehrer (1991) for a study of this phenomenon). In our Cournot example, dependencies in the stochastic realizations of past market prices serve as a device correlating players' future beliefs and strategies.

The need to distinguish subjective from objective knowledge in social interaction is not new or unique to this paper. Our notion of subjective equilibrium has its roots already in Van Hayek (1937). He proposes that, at equilibrium, "the individual subjective sets of data correspond to the objective data, and. . .in consequence the expectations in which plans were based are born out by the facts." Since Van Huyck, other economists have advocated and used such subjective notions, see for example Hahn (1973).

Also the newer literature on game theory contains an increasing number of concepts reducing the objective-knowledge assumed by Nash, and moving in the direction of subjective equilibrium. Rationalizable equilibria, see Bernheim (1984) and Pearce (1987), and the more recent Rubinstein and Wolinsky (1990)

rationalizable conjectural equilibria, are such examples. The notions most closely related to the ones proposed here are by Battigalli (1987) (see also Battigalli and Guaitoli (1988) and Battigalli, Gilli and Molinari (1992)), the self-confirming equilibrium of Fudenberg and Levine (1993), and the earlier version of subjective-equilibrium proposed in Kalai and Lehrer (1993a). Our convergence result is closely related to earlier Bayesian learning papers, for example Jordan (1991), Kalai and Lehrer (1993b).

The Notions of subjective Nash and subjective correlated equilibria proposed here generalize the earlier concepts in several ways. First, unlike the model proposed here, the papers just cited assume that the game is known and uncertainty is restricted to opponents' choice of strategies. Second, while some of the earlier papers assumed perfect monitoring of opponent's actions, the current paper does not. As a consequence the resulting notion of subjective equilibrium is more general. Also, our formal presentation of the concepts is developed through a model of infinitely repeated imperfect-monitoring stochastic-outcome game, while the earlier notions were defined on different classes. For example, self-confirming equilibrium was developed for finitely repeated but general extensive form games. It turns out however, that all the notions involved are natural enough and as a result, the modification of the equilibrium concepts as we move from one class of games to another, is fairly straightforward. We illustrate the formulation of subjective games and equilibria for general extensive form games in Section 6.

In recent years, researchers have been making heavy use of Nash equilibria to predict outcomes of social strategic interaction. An excellent example is modern industrial organization theory. Such analysis is carried in two stages. First the analyst formulates an abstract game describing the real situation. Then, assuming

that the same game is formulated by all the players, he computes its Nash equilibria as the set possible outcomes.

More recently, however, researchers are more cautious in using the above approach. It is recognized that Nash analysis is sensitive to the specification of the game and, thus, its predictions are not robust when game formulation is subjective. The problem is especially severe since it involves compounded lack of robustness where players mutually rely on each other's correct formulation. For example, if all the players formulated the game correctly but all of them except player A thought that player A's model is different, it is likely to lead them to change their choice of strategies. Worse yet, even at higher levels, the fact that player A believes that another player, B, formulates differently, may cause a significant change of strategies on the part of A's opponents.

Under the subjective approach the robustness issue is less severe. Here, the choice of a strategy by a player does not rely on game specifications of others, but is allowed to depend on independent subjective primitives. Thus, the compounding effect of non-robustness due to mutual reliance on correct specifications is eliminated.

Naturally, eliminating the assumption of the common availability of objective knowledge results in a significant reduction in prediction power. This is seen by the fact that the set of subjective equilibrium of a given game is in general larger than the set of Nash equilibria. We do not view this as a serious loss since we do not think that there was a real prediction power in the first place, because of the robustness issue.

We feel that a better prediction power can be obtained under the subjective approach provided that the analyst collects more data about the subjective beliefs players hold. The body of this paper contains such preliminary illustrations. For

example, when players believe that they are too small to affect market prices, the resulting subjective equilibrium of a finite player Cournot game yields competitive production. In another example, dealing with a homogeneous product dynamic Cournot oligopoly game, correct individual subjective assessments of aggregate market demand imply that subjective equilibrium yields the same behavior as Cournot predicts.

The literature on learning in strategic interaction has exploded over the last few years. It includes too large a number of bounded and myopic models to list here, as well as a large number of rational learning papers. A very partial sample of recent related rational-learning models includes Crawford and Heller (1990), Monderer and Samet (1990), Nyarko (1991b), Vives (1992), Koutsougeras and Yannelis (1993), Goyal and Janssen (1993), and Fujiwara–Grew (1993). Blume and Easley (1992) and Jordan (1993) present excellent critical evaluations of this approach. Also, a growing literature on strategic rational learning concentrating on reputation and forgiveness aspects is emerging--see, for example, Cripps and Thomas (1991), Schmidt (1991), and Watson (1992). These directions are especially important since forgiving strategies invite experimentation, a phenomenon that may create a coincidence of subjective with objective equilibria.

The present paper is also a direct contribution to the literature on players who do not know their own utility functions, as in the case-based approach of Gilboa and Schmeidler (1992). We discuss this after we study the multi-arm bandit example.

Two interesting connections to explore are with subjective variants of the Mertens and Zamir (1985) hierarchies of rationality model, as in Nyarko (1991a) and El-Gamal (1992), and with a new literature on endogenous uncertainty in economics, as in Chichilinksy (1992) and Kurz (1994). It seems that there should be

close relationships between subjective optimization and equilibria to these other new directions.

A forthcoming paper, Part II, will study additional properties of subjective equilibria. Among other issues, it will study: (a) general conditions under which subjective and objective equilibria coincide, and (b) the effect of the discount factor on experimentation with forgiving strategies.

2. Examples and Intuition

It is obvious that a subjective equilibrium may give rise to drastically different outcomes than objective equilibrium. Even the well known repeated prisoners' dilemma game with myopic players may be "solved." For example, if each player believes that whenever he acts non-cooperatively he will be severely punished by an outside force, his best response is to repeatedly act cooperatively. Thus, the two players play the fully cooperative path as a response to their beliefs. Moreover, their beliefs are not contradicted, since neither ever acts non-cooperatively to find out that his fear of severe punishment was not founded.

Before we turn, however, to additional multi-person examples with less "dramatic" beliefs, we start with the well known one person multi-arm bandit problem (see Wittle (1982) for the general problem, and see Rothschild (1974), and Banks and Sundanam (1993) for more recent references and economic applications). It turns out to be a special, stationary case, of our general formulation. The need to distinguish between subjective and objective equilibria becomes very clear here.

Example 2.1 (A Two-arm Bandit Game): The player in each period $t = 1, 2, ...$, has to engage in one of two possible activities, L and R. (A special case where these activities represent handles of two different slot machines motivates the name of this

problem.) Each activity, L and R, has a stationary payoff distribution, $\prod_L$ and $\prod_R$, describing independent probabilities of realized payoffs when the corresponding activity is used. The player's goal is to maximize the expected present value of his total payoff, discounted by some fixed parameter. Clearly, the optimal objective solution is to repeatedly use the activity with the higher per-play expected value.

What makes the problem interesting is that the player may not know $\prod_L$, $\prod_R$, or both. Instead, as he plays he observes payoffs generated by these distributions according to the actions that he uses. So every time he chooses to use L he sees the resulting payoff generated by $\prod_L$, and the same for R. But in every period, before making his choice, he knows the full history of his past choices and resulting payoffs. In order to maximize his expected payoff, depending on his discount parameter and subjective beliefs, it may pay him to experiment with both activities in order to learn something about their payoff distribution. Clearly, higher discount factors, representing more patient players in our conventions, lead to more experimentation, even if some immediate payoffs may seem to be sacrificed. But the problem is difficult and the question of how much and how to experiment depends in a fairly complex way on the subjective beliefs. These are described by prior probability distributions on sets of possible payoff distributions associated with each activity.

Suppose, for our example, that activity L generates payoffs of \$0 or \$2 with equal probabilities, i.e., $\prod_L (0) = \prod_L (2) = .5$. Let's also assume that the player knows that. On the other hand, he does not know $\prod_R$ and assigns positive probabilities λ^G and λ^B ($\lambda^G + \lambda^B = 1$) to it being one of two possible distributions $\prod^G$ and $\prod^B$. The "good" distribution $\prod^G$ has $\prod^G (2) = .6$ and $\prod^G (0) = .4$, but the "bad" distribution has $\prod^B (2) = .4$ and $\prod^B (0) = .6$. The following scenarios give rise to equilibria, or lack of such, of different types.

Scenario 1: $\prod_R = \prod^B$, λ^B is very high, and the player chooses to play repeatedly activity L. This is an objectively optimal strategy, since he chooses the optimal strategy against the true payoff distributions of the two machines. It is also subjectively optimal since, for sufficiently high λ^B, the best response is not to experiment and to just use activity L. Notice also that the beliefs of the players will remain the same throughout the play, since the only uncertainty is regarding $\prod_R$ but R is never used. In particular, the belief-confirmation property is satisfied. That is, given his strategy, his assessment of probabilities of his actual future payoffs is accurate. So his strategy with beliefs constitute also a subjective equilibrium. This is despite the fact that his conjectures, regarding hypothetical payoffs under different strategies, are not accurate.

Scenario 2: As before, λ^B is very high and the player uses repeatedly activity L, but now the real payoff distribution $\prod_R = \prod^G$. The player is best responding to his subjective beliefs, described by the high value of λ^B. Moreover, since he always uses activity L, he will never find out that his beliefs are very far from the truth. In this scenario we are at a subjective equilibrium, which is not an objective equilibrium. If the player knew that $\prod_R = \prod^G$ he would not want to stay with the constant left strategy.

Scenario 3: λ^G is high, the player uses repeatedly activity R, but $\prod_R = \prod^B$. This is obviously not an objectively optimal solution. But also subjective equilibrium fails. While the player maximizes initially against his beliefs, with increasingly high probability he will find out that his subjective beliefs are wrong, i.e., his posterior beliefs on $\prod^B$ will converge to 1, and as a consequence would not stay with the repeated use of activity R. In particular, belief-confirmation is violated here. The

belief that λ^G is high, together with the choice of always playing R, leads the player to optimistic assessment regarding his realized future payoffs. However, in the language of Van Huyck, his expectations will not be "born out by the facts" as he keeps playing R.

The previous example with the three scenarios illustrates the relationships of the different equilibria. Every objective equilibrium is a subjective one, when the subjective assessments happen to coincide with the true distributions. Because then, subjective and objective optimization are the same and belief confirmation is unavoidable. However, when the subjective assessments are not accurate, as in scenario 2, we may have a discrepancy. Thus, the set of subjective equilibria is really larger. But, as scenario 3 illustrates, not all strategies and beliefs constitute subjective equilibria.

In scenario 2 above, the discrepancy between subjective and objective equilibria is due to the fact that the player does not "know the game" he is playing. In this example, he does not know the payoff rules. When we move to multi-player situations, different types of information imperfections may cause such discrepancies. In the next example, even though both players fully know the game, imperfect monitoring of each other's actions brings about equilibria which are subjective but not objective. We refer the reader to Fudenberg and Kreps (1988) and Fudenberg and Levine (1993) for similar earlier examples.

Example 2.2 (Acting in the Dark): This symmetric 2-person game has two actions for each player: r for rest and a for act.

	r	a
r	0, 0	0, 1
a	1, 0	-1, -1

A player choosing r is paid 0 regardless of his opponent choice. A player that chooses a, on the other hand, is paid 1 if his opponent chooses r, but -1 if his opponent chooses a too.

We assume that the two players choose their actions repeatedly and simultaneously in the beginning of periods t = 1, 2, ... However, in each period, after the choices are made, each is only told his payoff, and is not told his opponent's choice. This means that, when he chooses to rest, he learns nothing about his opponent's choice. But when he chooses to act he learns, indirectly through his payoff, his opponent's choice.

We also assume that the players know all the information given above, i.e., they have common knowledge of the game. The only uncertainty each faces is regarding his opponent strategy.

Let A be the constant strategy of acting in each period and R be the constant rest strategy. It is easy to see that (A, R) and (R, A) are objective Nash equilibria of the repeated game with imperfect monitoring. These are equilibria because the best reply to A is R and vice versa.

What about (R, R)? The first player may be playing R because he thinks that player two is playing A. With the imperfect monitoring he never finds out that he is wrong and playing R against the conjecture that the other is playing A is as justified as playing R when the other player really plays A. So again we have a situation where each player chooses a strategy, R in this case, which is a best response to his

conjecture, that his opponent plays A, and what he observes does not contradict his conjectures. In other words (R, R) is a subjective equilibrium, even though it is not an objective Nash equilibrium.

Our next example is of a two player game. It illustrates several important items. First, moving to correlation, it shows a subjective correlated equilibrium which is not an objective correlated equilibrium of the same game. Second, it illustrates the process of learning and converging, in one step here, to the subjective correlated equilibrium. Finally, it illustrates the generality of a class of games we allow in our model. In particular, our stage game can be viewed as a multi-person, multi-arm bandit problem.

Example 2.3 (Winners and Losers Acting in the Dark): As in the previous example we consider a two player game with each having to repeatedly choose between resting (r) or acting (a), and each being informed only about his resulting payoffs. Again, a player that rests receives a zero payoff and a player that acts, at a period where his opponent rests, receives a payoff of 1. The above information is known to both players. However, now when both players act, a random pair of payoffs will be generated according to a fixed probability distribution $\prod_{a,a}$.

	r	a
r	0, 0	0, 1
a	1, 0	$\prod_{a,a}$

We consider different scenarios that may arise depending on the beliefs and actual payoff when both players choose to act.

We first restrict ourselves to the case where $\prod_{a,a}$ can take on only two

possible values: $\prod_{a,a} = \prod^{W,L}$ or $\prod_{a,a} = \prod^{L,W}$, defined as follows. $\prod^{W,L}(10, -1) = .99$ and $\prod^{W,L}(-1, 10) = .01$. In other words, under $\prod^{W,L}$ player 2 is most likely to lose while player 1 is most likely to greatly enjoy "winning" the conflict with player 2. Symmetrically, we define $\prod^{L,W}(10, -1) = .01$ and $\prod^{L,W}(-1, 10) = .99$.

Scenario 1: A standard common knowledge game. Nature moves first and chooses randomly with equal probability $\prod_{a,a}$ to equal $\prod^{W,L}$ or $\prod^{L,W}$. The realized choice, which is to be fixed now for the duration of the infinite game, is not revealed to the players. However, following standard game theoretic analysis, we assume that all the information above is common knowledge.

If the players are sufficiently patient, then each would want to learn if he is a "frequent winner" or a "frequent loser" in order to continue playing the game optimally. A reasonable Nash equilibrium of this Bayesian game has each player experimenting by acting at the first stage. If he loses he stops acting forever, but if he wins 1 or 10's he acts again. (After a while the computation of equilibrium becomes more complicated, since every time that he receives a 10 or a -1 he can update his prior as to the underlying $\prod_{a,a}$ being $\prod^{W,L}$ or $\prod^{L,W}$. Like the one-arm bandit problem, this is a relatively simple analysis. Once a player decides to rest at some stage he receives no new information. Assuming therefore a "once-rest, rest-forever" strategy, facilitates the computation of relatively simple equilibrium. We choose not to complete this computation here, since it is tangential to the points we wish to make.)

Scenario 2: Where both players are wrong and learn in one step to play a subjective correlated equilibrium which is not objective. Suppose the players believe everything

as in Scenario 1 and, therefore, choose Nash equilibrium strategies of the type described there. But assume that they are both wrong in that the payoff distribution of both acting, $\prod_{a,a}$, is really the random distribution $\prod^R$, defined as follows. $\prod^R (10, -1) = \prod^R (-1, 10) = \prod^R (-1, -1) = 1/3$. In other words, in each period when they both act it is equally likely, independently of the past, that one would lose, the other will win a lot, but also that they both lose.

Since both players choose to act in the first period, they will be paid according to a random draw of $\prod^R$. Therefore, there are three possible developments from period two on. With 1/3 probability, the first draw is a (10, -1). Following this they will each assign high subjective probability to $\prod_{a,a} = \prod^{W,L}$, and will continue playing the constant strategies (A, R). Similarly, with 1/3 probability they will be paid $(-1, 10)$, assign high probability to $\prod^{L,W}$, and play (R, A). But also with 1/3 probability they will draw (-1, -1) in the first period. This will lead each player to assign high probability to the distribution in which he is a loser and as a result the pair of constant strategies (R, R) will be played.

So if we consider the game, starting from period two on, we have a correlated strategy, assigning probabilities 1/3 to (R, A), to (A, R) and to (R, R). Correspondingly, we have correlated beliefs where, with probability of 1/3 each, the players respectively assign high subjective likelihoods to the payoff distribution being $(\prod^{L,W}, \prod^{L,W})$, $(\prod^{W,L}, \prod^{W,L})$ and $(\prod^{L,W}, \prod^{W,L})$. This is a subjective correlated equilibrium, since after each period 1 outcome, the correlated strategies are best response to the correlated beliefs and the induced subjective distributions on the future play of the game coincide with the objective one.

Consider, for example, the initial message to be the draw (-1, -1). Now each player believes that he has encountered an acting opponent in the first period. Moreover, given that he lost (he does not, of course, even consider the possibility

that his opponent lost too) his updated posterior beliefs are that he is very likely a frequent loser and he decides to stay out forever. Since both stay out forever, their beliefs regarding their future payoffs in the game are accurate.

Since the actual expected payoffs of the action vector (a, a) are (2.66, 2.66), the correlated strategies 1/3 on (A, R), 1/3 on (R, A) and 1/3 (R, R) are not a correlated equilibrium of the real repeated game.

3. Subjective Equilibrium of a Single Decision Maker

We consider a player with a nonempty finite set of actions A, a countable set of outcomes (consequences) C, and bounded a von Neumann-Morgenstern utility function $u: A \times C \rightarrow \mathbf{R}$.

Dynamically, the player will choose actions $a^1, a^2, \ldots$ from A. In every period t, after he chooses the action a^t, an outcome $c^t \in C$ will be stochastically determined, reported to him, and he will collect the payoff $u(a^t, c^t)$. The player's objective is to maximize the present value of his expected utility discounted by a fixed parameter λ, $0 < \lambda < 1$.

The above formulation implicitly assumes that the player knows A, C, u and λ. What he does not know is the stochastic rule by which outcomes are generated.

Examples of such problems are numerous. We will analyze the multi-arm bandit problem, where A represents a set of possible "arms" or activities to use, $c \in C$ represents a stochastically generated payoff, and $u(a,c) = c$. The stochastic choice of the outcome c in this example will be stationary and its distribution will depend entirely on the chosen a.

A more complex economic example concerns a producer in an oligopoly whose action a^t in each period t describes a chosen production level. Here, an outcome c^t describes his resulting market price. The stochastic determination of c^t

is a function of his production level a^t, the production choices of his competitors, and a demand function which depends on the joint production vector plus a random noise. Here we will not assume stationary determination of outcomes (prices) since the competitors are likely to change their production levels as they too observe the behavior of the market.

In the general formulation, the determination of outcomes is described by a stochastic environment response function denoted by e. For every history of actions and outcomes, $h^t = (a^1,c^1,...,a^t,c^t)$, and for any t + 1 period action a^{t+1}, e defines a probability distribution over C. Formally, $e|_{h^t a^{t+1}}(c)$ denotes the probability that the outcome c will be chosen after the play consisting of the history h^t followed by the action a^{t+1}. Thus the above values must be nonnegative and sum to 1 over the possible values of c for any fixed h^t and a^{t+1}. The unique empty history h^0 is allowed and thus $e|_{h^0 a^1}$ describes the distribution of initial outcomes as a function of every chosen initial action a^1. (When it does not create confusion, to simplify notation we will omit some time-superscripts, e.g., write $e|_{ha}(c)$).

If the player knows the environment response function e, his problem is to choose a (behavior) strategy f to maximize the present value of his expected payoff computed with the distribution generated by his strategy and e. Formally such a strategy f assigns a probability distribution over the action set A for every history of past actions and outcomes. Thus, $f|_{h^t}(a)$ represents the probability that action a will be chosen in period t + 1 if the player observed the history h^t. Fixing h^t, $f|_{h^t}(a)$ must sum to one as we vary $a \in A$.

We choose not to restrict our analysis to pure strategies, where each $f|_{h^t}$ assigns probability one to a single $a \in A$. Such a restriction, even if not significant for the one player case, would limit the scope of the analysis in the sections that follow.

To consider the expected present value of utility resulting from a strategy f, we first must describe the underlying probability space. It consists of a set Z of infinite play paths of the form $z = (a^1,c^1,a^2,c^2,...)$. For a history h^t, as described above, we will abuse notation and let is also denote the cylinder set in Z, consisting of all infinite play paths z whose initial t−period segment coincides with h^t. As usual, the σ−algebra used for Z is the one generated by all cylinder sets h^t and to specify a probability on Z it suffices to assign consistent probabilities to all h^t's.

We do this inductively in the usual way. Given a strategy f and an environment reaction function e, we define $\mu_{f,e}(h^0) = 1$. For h^{t+1} described by h^t followed by a^{t+1}, c^{t+1}, we define $\mu_{f,e}(h^{t+1}) = \mu_{f,e}(h^t)f|_{h^t}(a^{t+1})e|_{h^t a^{t+1}}(c^{t+1})$.

Now we can define utility functions for strategies. First, the utility assigned to a play path $z = (a^1,c^1,a^2,c^2,...)$ is computed by $u(z) = \Sigma\, \lambda^{t-1}u(a^t,c^t)$. The utility of a strategy f and an environment reaction function e is computed to be $u(f,e) = \int u(z)d\mu_{f,e}(z)$.

As stated earlier, the player's objective is to choose f that maximizes u(f,e). However, since we assume that the player does not know e, he cannot solve the above problem.

Taking a subjective approach, we assume that the player holds an endogenously given subjective belief about the environment reaction function, $\tilde{e}$, and that he chooses f to maximize $u(f,\tilde{e})$. But we do not assume that $\tilde{e}$ coincides with e. When this is the case we say that f is subjectively optimal relative to $\tilde{e}$. If f is optimal relative to the "real" e we say that it is objectively optimal or just optimal.

Remark 3.1

A. Beliefs Over a Set of Possible Environments. While in the above formulation the player's subjective belief is restricted to be a single environment

reaction function it is really more general. For example, if the player assigned prior probabilities $q_1,q_2,...,q_n$ to a set of possible environment reaction functions $\tilde{e}_1,...,\tilde{e}_n$, he could replace this belief system by a single equivalent belief function $\tilde{e}$. This is done using the usual Bayes updating construction as, for example, in Kuhn's (1953) theorem. After every history h^t one computes posterior probabilities $\tilde{q}_1,...,\tilde{q}_n$ for the environments $\tilde{e}_1,...,\tilde{e}_n$ and assign probabilities to the next outcome according to the $\tilde{e}_i$'s weighted with the updated posteriors. We do such a construction in our example of a multi-arm bandit problem discussed later.

B. Imperfect Updating of Environmental Reactions. Updating posterior beliefs, as described above, assumes a type of consistency and perfect rationality on the beliefs of the player. However, the abstract formulation described by a single $\tilde{e}$, which is a function that can be freely defined after every history, allows for more general and imperfect updating. For example, a player with Bayesian posterior probabilities, $\tilde{q}_1,...,\tilde{q}_n$, can adjust some up and some down if he choose to put less weight than the correct one on small probability posteriors.

The discrepancy between the real environment response function, e, and the subjective one, $\tilde{e}$, may make the player alert to the fact that his assessment is wrong. Given his choice of strategy f, his assessment of the stochastic evolution of his future outcomes is given by $\mu_{f,\tilde{e}}$, while the real evolution follows the distribution $\mu_{f,e}$. If, however, $\mu_{f,\tilde{e}} = \mu_{f,e}$ then it is impossible for him to detect, even with sophisticated statistical tests, that he is wrong. This is despite the fact that serious discrepancies may exist between e and $\tilde{e}$. These discrepancies, however, are non-observable under his chosen strategy. With such discrepancies, even if f is subjectively optimal it may be objectively suboptimal but the player could not determine that, and will have no cause to change his assessment or his strategy. This gives rise to the following

definition.

Definition 3.1: The strategy f with the environment reaction function e is a subjective equilibrium relative to the belief $\tilde{e}$ if the following two conditions hold.

1. Subjective-Optimization: f maximizes $u(f,\tilde{e})$, and
2. Belief-Confirmation: $\mu_{f,\tilde{e}} = \mu_{f,e}$.

Remark 3.2 (Optimizing Implies Experimenting): Reflecting on the definition above, a subjective equilibrium can be suboptimal because, and only because, its assessment of outcome probabilities off the equilibrium play path is wrong. An obvious remedy to such a deficiency is for the player to experiment, in order to learn to the greatest extent possible, the off-path outcome probabilities. When and how much to experiment are difficult questions. While under-experimentation may be suboptimal, over-experimentation may also be so. Computing the optimal level of experimentation requires knowledge of real distributions, which the player does not possess. However, under the subjective approach, it is naturally incorporated into his subjective optimization problem.

Consider, for example, a two-arm bandit player, with two competing activities, L and R, of Example 2.1. Suppose each activity has a stationary payoff distribution Π_L and Π_R. Assume for simplicity, as we did there, that the subjective beliefs are accurate on left, $\tilde{\Pi}_L = \Pi_L$, with expected utility of 1 for every use of L. On the other hand, for R, the player believes that there are the two distributions Π^B and Π^G, one of which was drawn initially with probabilities .90 and .10, respectively. Recall that the corresponding expected values are 0.8 and 1.2. By the law of large numbers, sufficiently long use of R will reveal to the player whether Π^G or Π^B was

drawn. Depending on his discount parameter, his subjective optimization will determine the optimal experimentation strategy. If the future is important enough, the ten percent chance that eventual generation of a payoff stream with expected value of 1.2 in each period will dictate an initial experimentation period. But if future payoffs are sufficiently unimportant, it would be subjectively suboptimal to experiment.

The optimal strategy in the definition of subjective equilibrium above already includes a subjectively optimal level of experimentation. The actual computation of such optimal strategies is done using the well known Gittins index, see Wittle (1982).

We will see in the sequel that under a certain condition, relating the belief to the truth, a subjective optimizer must converge eventually to a subjective equilibrium. In any finite time, however, he may converge only to an $\epsilon-$subjective equilibrium where the subjective distribution, $\mu_{f,\tilde{e}}$, is only close to the objective one, $\mu_{f,e}$. To make this precise we must first discuss notions of closeness of distributions.

Definition 3.2: For a given $\epsilon > 0$ and two probability distributions, μ and $\tilde{\mu}$, we say that $\tilde{\mu}$ is $\epsilon-$close to μ if for any event A, $|\mu(A) - \tilde{\mu}(A)| \leq \epsilon$.

Remark 3.3: Interpretations of Closeness of Distributions. We say that $\tilde{\mu}$ is $\epsilon-$near to μ if there is an event Q, with $\mu(Q)$ and $\tilde{\mu}(Q) \geq 1 - \epsilon$, satisfying $|1 - \mu(A)/\tilde{\mu}(A)| \leq \epsilon$ for every event $A \subseteq Q$ (we assume in the above that $0/0 = 1$).

As was shown in Kalai and Lehrer (1993c), the two notions, $\epsilon-$closeness and $\epsilon-$nearness are asymptotically equivalent, i.e., by making the distributions sufficiently close in one sense, we can force them to be as close as we wish in the other sense. Thus, limit results, where we obtain eventual arbitrary closeness of two

measures, are the same in both senses.

While the notion of ϵ-closeness is easier to state, the notion of $\epsilon-$nearness is more revealing. First notice that $\epsilon-$closeness says little on small probability events. For example, we can have $\tilde{\mu}(A) = 2\mu(A)$ and still have $\tilde{\mu}$ be $\epsilon-$close to μ provided that $\mu(A) < \epsilon/2$. On the other hand, $\epsilon-$nearness shows that this can be the case but not on events $A \subseteq Q$. Within the large set Q the ratios of the probabilities must be close to 1. This has important implications for conditional probabilities, which take on special importance in models with infinite horizons.

Recall that our discussion of closeness of the measures $\tilde{\mu}$ and μ is motivated to capture the idea that a player believing $\tilde{\mu}$ but observing events generated by μ is not likely to suspect that $\tilde{\mu}$ is wrong. The notion of ϵ-closeness captures this idea for large events. Our player, however, after a long play is likely to observe small probability events consisting of long chains of events. His forecast of future events then will be obtained by assigning probability to future events conditional on having observed low probability events. Thus, if our notion of closeness of $\tilde{\mu}$ and μ are such that the conditional probabilities they generate remain close, then the player using $\tilde{\mu}$ is not likely to suspect his $\tilde{\mu}$ even in the far future.

ϵ-nearness, and thus its asymptotically equivalent notion of ϵ-closeness, fulfills this property to a large extent. Since

$$\frac{\mu(A|B)}{\tilde{\mu}(A|B)} = \frac{\mu(A \text{ and } B)}{\tilde{\mu}(A \text{ and } B)} \frac{\tilde{\mu}(B)}{\mu(B)},$$

we can deduce that if A and B are events in Q, no matter how small, then closeness to 1 of the two factors in the right side implies closeness of the conditional probabilities in the left side.

Definition 3.3: Given $\epsilon > 0$, a strategy f, and environments e and $\tilde{e}$, we say that f is an ϵ-subjective equilibrium in the environment e relative to $\tilde{e}$ if the following two conditions hold:

1. Subjective Optimization: f maximizes $u(f,\tilde{e})$, and
2. ϵ-Belief-Confirmation: $\mu_{f,\tilde{e}}$ is ϵ-close to $\mu_{f,e}$.

Convergence of a subjectively optimal strategy to a subjective equilibrium is not guaranteed in general but is true under sufficient conditions of compatibility of the beliefs with the truth. The relationships between notions of compatibility, notions of convergence, and alternative notions of ϵ-subjective equilibrium, involve detailed mathematical analysis. To proceed with the presentation of the subjective approach, we present one such notion of compatibility that works well with our notion of ϵ-closeness (or ϵ-nearness) as defined above. For alternative concepts we refer the reader to Lehrer and Smorodinsky (1993).

Definition 3.4: We say that the subjective evolution described by $(f,\tilde{e})$ is compatible with the one generated by (f,e) if the distribution $\mu_{f,e}$ is absolutely continuous with respect to $\mu_{f,\tilde{e}}$, $\mu_{f,\tilde{e}} \succ \mu_{f,e}$. This means that for every event A,

$$\mu_{f,e}(A) > 0 \Rightarrow \mu_{f,\tilde{e}}(A) > 0.$$

In other words, events considered impossible according to the subjective belief of the agent, i.e., having subjective probability zero, are really impossible, i.e., they have objective zero probability.

Our goal is to show that after a sufficiently long time T, a subjective optimizer will play essentially an ϵ-subjective equilibrium for arbitrarily small ϵ. To make this

formal we need to describe the environment response functions and strategies induced on the "new" problem starting from time T on.

Definition 3.5: Let e be an environment response function, f a strategy, and h a history of length t. Define the environment response function e_h and the strategy f_h induced by h by

$$e_h|_{\bar{h}}(c) = e|_{h\bar{h}}(c) \text{ and}$$
$$f_h|_{\bar{h}}(a) = f|_{h\bar{h}}(a).$$

The notation $h\bar{h}$ used above denotes the concatenation of the histories h and $\bar{h}$, i.e., the history whose length is the sum of the lengths of h and $\bar{h}$ obtained by starting with the elements of h and continuing with the elements of $\bar{h}$.

Theorem 3.1: Let f be a subjectively optimal strategy relative to $\tilde{e}$ in the environment e, and assume that $(f,\tilde{e})$ is compatible with (f,e). For every $\epsilon > 0$ there is a time T such that with probability greater than $1 - \epsilon$, f_h is an ϵ-subjective equilibrium in the environment e_h relative to the beliefs $\tilde{e}_h$ for every history h with length greater than T.

The probability $1 - \epsilon$ in the statement of the theorem is the objective one, computed by $\mu_{f,e}$.

The proof of the theorem follows immediately from the seminal "merging of opinions" theorem in Blackwell and Dubins (1962) (see also Kalai and Lehrer (1993c) for extensions and alternative statements).

Example 3.1. The Multi-arm Bandit Problem

In the general formulation, we think of A as any finite set of activities that can be used repeatedly in periods t = 1,2,... . A countable set of outcomes C consists of real numbers representing possible payoffs. For each activity $a \in A$ there is a fixed probability distribution Π_a over C with $\Pi_a(c)$ describing the (past independent) probability of the outcome c being realized when the action a is taken. The player's goal is to choose a sequence of actions $a^1,a^2,...$, with each $a^t \in A$, that will maximize the present value of his expected payoff. However, he does not know the distributions, Π_a's, and whenever he uses the action a^t at time t, he is told his realized payoff, which was drawn according to Π_{a^t}. Naturally, he can use this and all previous information before making his next choice, a^{t+1}.

In our general formulation, this example is modeled with A and C being described as above, u(a,c) = c, and a stationary environment function $e|_{ha}(c) = \Pi_a(c)$. Our player, not knowing the functions Π but knowing the stationary structure of the model, assumes that for every a, the distribution Π_a was chosen from among m possible distributions $\Pi_a^1,...,\Pi_a^m$ with positive prior probabilities $\lambda_a^1,...,\lambda_a^m$. We assume that Π_a indeed equals Π_a^j for some j.

The subjective environment response function, $\tilde{e}$, is computed by the standard method of Bayesian updating. First we compute inductively posterior probabilities $\lambda_a^j|_h$, j = 1,...,m, for every a and h. Initially, $\lambda_a^j|_{h^0} = \lambda_a^j$. And for a history of the form $\bar{h}$ obtained by concatenating a history h with an action outcome pair $(\bar{a},c)$ $\lambda_a^j|_{\bar{h}} = \lambda_a^j|_h$ if $\bar{a} \neq a$, and $\lambda_a^j|_{\bar{h}} = \lambda_a^j|_h \, \Pi_a^j(c) / \left[\sum_i \lambda_a^i|_h \, \Pi_a^i(c)\right]$ if $\bar{a} = a$. Then $\tilde{e}$ is defined by $\tilde{e}|_{h,a}(c) = \Sigma_i \, \lambda_a^i|_h \, \Pi_a^i(c)$.

Since we assumed above that Π_a is assigned positive prior probability, for every strategy f, $(f,\tilde{e})$ is compatible with (f,e). Thus by Theorem 3.1 for every $\epsilon > 0$ we can find a large enough time T such that with probability of at least $1 - \epsilon$ the strategy and beliefs of the player from time T on constitute an ϵ-subjective

equilibrium.

Corollary 3.1: Suppose that the activities are strictly ranked by expected value, i.e., distinct objective expected values are generated by distinct activities, then for every $\epsilon > 0$ there is a time t such that with probability greater than $1 - \epsilon$ the subjectively optimizing player described above will use only one activity from time t on.

Proof: We may assume without loss of generality that σ is pure. (If σ is a subjectively optimal behavior strategy, we take a look at any pure strategy in the support of σ.) We show that with probability 1 there is a (random) time t from which on σ prescribes playing one arm only. This certainly implies the corollary.

Assume to the contrary that there exists an event, R, with positive probability such that on every infinite history $h \in R$ there are infinitely many truncations of h,h^t, after which σ uses at least two arms. We denote by σ_{h^t} the continuation of σ after the finite history h^t.

From Theorem 3.1, we deduce that on almost every $h \in R$, σ_{h^t} is a δ_t−subjective equilibrium, where $\delta_t \to 0$. We take one $h \in R$ and consider the sequence of times t such that σ_{h^t} prescribes the arm a_1 first and $\sigma_{h^{t+1}}$ prescribes the arm a_2 ($a_1 \neq a_2$) first. We proceed by the following lemma to the contradiction.

Lemma 1: Let σ_t be a δ_t-subjective equilibrium, where $\delta_t \to 0$, then any limit of $\sigma_t (t \to \infty)$ is a subjective equilibrium.

Proof: Clearly, every limit of σ_t is optimal against the limit of the corresponding beliefs and moreover, confirms this limit belief. //

Lemma 2: If $\bar{\sigma}$ is a subjective equilibrium in the set -up of Corollary 3.1, it uses only one arm, with probability 1.

Proof: Let A′ be the set of those arms used with a positive probability under $\bar{\sigma}$. Since $\bar{\sigma}$ is subjectively optimal in the grand game (with the full set of arms, A), it is also subjectively optimal in the reduced game (with A′ only). As $\bar{\sigma}$ is subjective equilibrium with A it is an objective equilibrium with A′ (simply because there is a full knowledge about the expected payoffs of all the arms available, A′).

However, as an objective equilibrium, $\bar{\sigma}$ should prescribe using only one arm, the best one. //

Returning to the proof of the corollary, recall that σ is pure, that σ_{h^t} prescribes the arm a_1 and that $\sigma_{h^{t+1}}$ prescribes the arm a_2. Denote by w^{t+1} the outcome that forms with the history h^t, the longer history, h^{t+1}. (I.e., h^{t+1} is the concatenation of h^t and w^{t+1}.) Since there exist only finitely many w^{t+1} and infinitely many t, we may assume that all the w^{t+1} are the same. As the probability to get w^{t+1} by using the arm a_1 is stationary, say, $p > 0$, we deduce that σ_{h^t} prescribes using the arm a_2 at the second stage with probability p. Therefore, any limit of σ_t, $\bar{\sigma}$, assigns to two arms a_1 and to a_2 positive probabilities. By Lemma 1, $\bar{\sigma}$ is a subjective equilibrium which contradicts Lemma 2. //

Remark 3.4 (Players who do not know their own utilities): Learning one's own utility function is an important problem in decision theory--see, for example, Gilboa and Schmeidler (1992) for a new approach and recent references. It deals with a player that can choose repeatedly activities a from a set A but does not know his own utility function u(a). Our general formulation assumes that the player has a

known utility function, u(a,c), defined on actions and their consequence. However, it includes the case of not knowing a function u(a) as a special case. To illustrate this point, observe that not knowing your own utility can be viewed as a special case of the multi-arm bandit problem with the set C representing numerical payoffs, and the unknown $\Pi_a(c)$ assigning probability one to $c = u(a)$.

4. Multi-Person Subjective Equilibria

4.1 The Repeated Stochastic-Outcome Game

We now assume that there are n−players, $n \geq 1$, each having a finite set of actions A_i, a countable set of outcomes C_i, a utility function $u_i: A_i \times C_i \rightarrow \mathbf{R}$, and a discount parameter λ_i. Also, as before, each players knows his individual data above, and would like to choose a sequence of actions, $a^1, a^2, \ldots$, to maximize the present value of his expected utility. But, again, he does not know the rule of how his actions affect his outcomes, i.e., his environment response function.

Taking the approach of the previous section, we could assume that each individual starts with a subjective belief about his environment, described by an $\tilde{e}_i$, chooses an optimal strategy f_i relative to $\tilde{e}_i$, and conclude that eventually each player will play a subjective equilibrium. However, we are now interested in the long term interactive equilibrium behavior of the players, and for that purpose we must first be more explicit about how the actions of one player enter the environment function of another.

We describe these cross affects by a collection of probability distributions, Π_a, defined for every action vector $a \in A \equiv \times_i A_i$. More precisely, $\Pi_a(c)$ denotes the probability that the outcome vector $c \in C \equiv \times_i C_i$ be realized if the vector of actions a is taken. Thus, for a fixed i, the above quantities must sum to 1 as we vary c. Notice that the distributions Π_a, together with the action sets A_i and the utility

function u_i fully determine an n−person stage game, G. In this game, for every action vector a, player i's (expected) utility is computed to be $u_i(a) = \Sigma_c$ $u_i(a_i,c_i)\Pi_a(c)$. We refer to such a game as a stochastic-outcome game.

The above game will be played repeatedly as follows. In every period $t = 1,2,...$ each player, being informed of his past actions and realized outcomes, will choose action $a_i^t \in A_i$. Then, based on the vector of choices, a^t, nature will choose a vector of outcomes $c^t \in C$ according to the distribution Π_a. Player i will be informed of his own outcome, c_i^t, will collect the payoff $u_i(a_i^t,c_i^t)$, and will proceed to choose a_i^{t+1}, and so on. Overall individual payoffs will be computed to be the present value of the total expected utility discounted by the individual discount parameters, λ_i. We denote the infinitely repeated game described above by G^∞.

Example 4.1.1: A Cournot Game with Differentiated Products. We assume that each of the n players is a producer of a certain good, with A_i denoting the set of his possible period production levels. Now C_i describes a set of period market prices producer i may realize. Thus, for every vector of production levels $a \in A$, $\Pi_a(p)$ describes the probability of the vector of individual prices $p = (p_1,...,p_n)$ being realized. The utility of player i is defined as usual by his resulting revenue minus cost, $a_ip_i - g_i(a_i)$, with g_i denoting his production cost function. Thus, in each period the player knows his previous production levels and prices, and based on this knowledge he chooses his next production level.

When all producers produce a homogeneous product, and face the same market price, we model the situation by restricting the support of each Π_a to p's with $p_1 = p_2 = ... = p_n$.

Remark 4.1.1: Imperfect Versus Perfect Monitoring. While our general

formulation, with each player being informed only of his own realized actions and outcomes, describes imperfect monitoring and other types of information imperfection, it includes as special cases games with more monitoring and common information. For example, perfect monitoring in the Cournot example above could be specified by letting each player's reported outcome, $c_i = (a_1,...,a_n,p_i)$. So the outcome reported to player i includes all the production levels but only his realized price. Full common knowledge of histories can be modeled by letting individually reported outcomes include all production levels and all realized prices, i.e., $c_i = (a_1,...,a_n, p_1,...,p_n)$.

Regardless of how the c_i's and Π are defined, however, under the convention that a player knows all his previous realized actions and outcomes before choosing his next action, our games always have perfect recall in Kuhn's (1953) sense.

Our general subjective approach will assume that there is a real "objective game," G^∞ as defined above, being played. We will depart, however, from the traditional game theory assumption that the players know the game. Instead we will define the notion of a subjective game to include the objective game and the beliefs of the individual players. Such subjective beliefs will be modeled by subjective environment response functions as defined in the previous section. It will ease the exposition, however, if we first review and establish the notations needed for the objective notions of Nash and correlated equilibria.

Formally, we define a history of length t, h^t, to consist of a vector $(a^1,c^1,...,a^t,c^t)$ where each $a^j \in A$ and $c^t \in C$. An individual player history $h_i^t = (a_i^1,c_i^1,...,a_i^t,c_i^t)$ with each $a_i^j \in A_i$ and $c_i^j \in C_i$. A play path $z = (a^1,c^1,a^2, c^2,...)$ and it induces finite histories h^t and finite individual histories h_i^t by taking projections to the first t elements and then taking projections to the i−th component.

A strategy of player i is a function f_i describing the probability that he takes a specified action after a specific history. Formally, $f_i|_{h_i}(a_i)$ denotes the probability that he would choose action a_i after observing his individual history h_i.

Following standard game theory, one defines the utility function $u_i(f)$ for every strategy vector $f = (f_1,...,f_n)$, and a Nash equilibrium to be a vector f^* with each f_i^* maximizing $u_i(f^*_{-i},f_i)$. (Here and elsewhere, f_{-i} denotes a vector of strategies of all players but i where (f^*_{-i},f_i) denotes the vector where all players but i play their star strategy but i plays f_i.) To define the (expected) utility functions one needs to first establish the probability space describing the possible plays of the game.

We let Z denote the set of (infinite) play paths, and as before we let h^t denote a history of a finite length t but also the cylinder set defined by it. Given a strategy vector f we define the probability distribution it induces on finite histories, μ_f, inductively. For the empty history $\mu_f(h^0) = 1$, and assuming that μ_f was defined for all histories of length t, we define it for histories h of length t + 1 by

$$\mu_f(h,a,c) = \mu_f(h) \times_i f_i|_{h_i}(a_i)\Pi_a(c).$$

Since the above construction defines consistent probabilities for all cylinder sets it defines the distribution μ_f on the set of play paths.

Now for every play path $z = (a^1,c^1,a^2,c^2,...)$ we define $u_i(z) = \sum_t \lambda_i^{t-1}u_i(a_i^t,c_i^t)$ and for a vector of strategies f we define $u_i(f) = \int u_i(z)d\mu_f(z)$.

Often the strategies of the players in the repeated game are correlated since their choice depends on correlated past individual messages. Formally, such a correlation device is described by two components. First is a nonempty countable set of message vectors, $M = \times_i M_i$, with each M_i denoting the set of player i's

messages. The second component is a probability distribution p defined on M.

A vector of correlated strategies, $f = (f_1,...,f_n)$ for the game G^∞, is defined by amending a correlation device to the beginning of the game. This is done by replacing the unique empty history by all possible elements $m \in M$ and allowing a player's strategy to depend on his reported initial message m_i. Formally, a history of "length zero" is now any element of M, history of length t is of the form $(m,a^1,c^1,...,a^t,c^t)$ and a play path $z = (m,a^1,c^1,a^2,c^2,...)$. Individual histories are described as before by projecting to the player's component. So an individual history of player i is a vector of the form $(m_i,a_i^1,c_i^1,...,a_i^t,c_i^t)$. Now a vector of correlated strategies $f = (f_1,...,f_n)$ has each f_i describe a distribution over player i's actions for every individual history with an initial individual message. In other words, it is a vector of standard behavior strategies for the game with the initial correlation device, the correlated game, (M,P,G^∞).

The utility of player i is computed as before to be his expected present value where the probability distribution on the expanded Z includes the initial distribution p. Thus we only need to modify the distribution over length zero histories by defining $\mu_f(m) = p(m)$. The probability of longer histories are defined inductively as before.

A vector of correlated strategies, f, is a correlated equilibrium of G^∞ if it is a Nash equilibrium of the correlated game (M,p,G^∞) as defined above, for some correlation device (M,p).

As in the previous section, we will be interested in the play of the repeated game starting after a long time T. In the "new" game correlation cannot be ruled out since each player strategy from time T on, may depend on his outcomes up to time T. And, in general, these outcomes are correlated.

Formally, given a vector of strategies for G^∞, f, and a positive integer T, we

define the induced vector of correlated strategies $f^T = (f_1^T,...f_n^T)$ as follows. M is the set of length T histories and p is the distribution μ_f restricted to M. Following a history consisting of an initial message m_i followed by h_i, f_i^T will randomize over A_i with the same distribution that f_i induced in the original game after the history obtained by concatenating m_i with h_i.

Remark 4.1.2: Nash Equilibrium Induces Correlated Equilibrium in Later Games. It is easy to see that if we start with a Nash equilibrium f, then f^T as defined above is only a correlated equilibrium of the repeated game starting at time T, see Lehrer (1991). Thus, in general games with imperfect monitoring, Nash equilibrium "deteriorates" to become only correlated equilibrium after time. This observation has important implications for learning theories. It suggests that, in general, we can at most hope to converge to correlated equilibrium.

Clearly, in the construction above, we could have started with a vector of correlated equilibrium for G^∞, to conclude that it induces a correlated equilibrium after any time T.

4.2 The Individual Environment Response Functions of the Repeated Game

In our stochastic-outcome games, to compute his best response strategy, a player does not have to know the game or his co-players' strategies. It suffices to know his one person decision problem, induced by the game and their strategies. This decision problem can be fully described by an environment response function as described in Section 3. For example, in the oligopoly game of the previous section, if a player knew his correct price distribution, after every history of play and for every one of his production levels, he would not need to have any information about his opponents and their strategies in order to compute his own optimal strategy.

The above observation will be especially useful in the next section, where we actually assume that the player does not know the game he is playing. But before we turn to the equivalent subjective concept, we first describe formally the objective notion.

We consider a repeated stochastic outcome game G^{∞} as in the previous section, and a fixed player i. As in Section 3, an environment response function for him, e_i, describes a probability distribution over his (next) outcomes after every history of play observed by him and an action chosen by him. More precisely, $e_i|_{h_i a_i}(c_i)$ is the probability of his next outcome being c_i after observing the individual history h_i and choosing the action a_i.

When the opponents' strategy vector, f_{-i}, is known, the computation of the induced environment function, e_i, is straightforward. For every history of length t, h_i, action a_i, and outcome c_i, we choose a strategy f_i for player i under which the individual history h_i followed by a_i has positive probability (or simply let player i play the actions of h_i up to time t, then a_i, and anything afterwards) and let μ_f be the induced distribution on play paths. Then define $e_i|_{h_i a_i}(c_i)$ to be the μ_f conditional probability of c_i being player i's outcome at time t + 1, given the individually observable play $h_i a_i$. If under the opponents' strategies, $h_i a_i$ is impossible, no matter what strategy player i chooses, then $e_i|_{h_i a_i}$ can be defined arbitrarily (since this situation will not arise).

Following the discussion above, it is straightforward to conclude the following equivalence.

Proposition 4.2.1: A vector of strategies f is a Nash equilibrium iff each player's strategy, f_i, is optimal relative to his environment function, e_i, induced by f_{-i}.

Clearly, the above discussion and definitions are also applicable to correlated versions of the repeated game, with an initial correlated device (M,p). In this case, as before, each zero length history consists of a message vector m and all other histories, individual or not, start with an initial message. Again, a correlated strategy vector f is a correlated equilibrium if and only if each f_i is a best response to the individual environment functions induced by f_{-i} (this is now in the game with initial correlation).

Equivalently, one can discuss these notions on the strategies induced by initial messages. For m_i, a positive probability message for player i, let f_{m_i} and e_{m_i} be his induced strategy and environment function after receiving the message m_i ($f_{m_i}|_{h_i} = f_i|_{m_i h_i}$ and $e_{m_i}|_{h_i a_i} = e_i|_{m_i h_i a_i}$).

Proposition 4.2.2: A vector of correlated strategies in the game (M,p,G^∞) is a correlated equilibrium if and only if for every player i and every positive probability message m_i, f_{m_i} is optimal relative to e_{m_i}.

Proposition 4.2.2 is identical to Proposition 4.2.1 with the exception that the optimization is checked only after the zero length histories. Since the zero length histories do not have any payoff, nor strategic choices, checking for optimal behavior after them involves no loss of generality.

4.3 The Subjective Game and Equilibrium

In this section we assume that a real game, G^∞ as defined before, will be played, but that the players do not necessarily have full knowledge of the game. We assume that they each know his own components, i.e., feasible actions, possible consequences, and utility functions. But we do not assume that each player knows

his opponents' possible strategies and utility functions. He may not even know who his opponents are and how many of them may be playing.

We model such a situation by assuming that each player holds a subjective belief about his environment described by an environment response function, $\tilde{e}_i$, as defined in the previous section. The player will choose a strategy f_i to be optimal relative to the subjective environment function $\tilde{e}_i$. These choices, made by all n-players, result in a vector of strategies $f = (f_1,...,f_n)$, which, in turn, induce objective environments $e_1,...,e_n$. As we already discussed in Section 3, there is no reason to assume that $e_i = \tilde{e}_i$ and if significant differences exist between the subjective and the objective distributions induced by them, the player will observe that his assessments are wrong, update his subjective beliefs, and modify his strategy.

However, an equilibrium situation can arise even if $\tilde{e}_i \neq e_i$, provided that the disagreements of the two functions are restricted to be after histories that are not observable, i.e., have zero marginal probabilities. When this is the case for each player, we are in a subjective equilibrium of the game.

To make this precise let μ_{f_i,e_i} and $\mu_{f_i,\tilde{e}_i}$ be, respectively, the objective and subjective distributions induced on player i's play paths.

Definition 4.3.1: Let $f = (f_1,...,f_n)$ be a vector of strategies, and $e = (e_1,...,e_n)$ be the induced environment functions. Let $\tilde{e} = (\tilde{e}_1,...,\tilde{e}_n)$ be a vector of subjective environment response functions. The pair $(f,\tilde{e})$ is a subjective Nash equilibrium of the game G^∞ if for each player i the following two conditions hold:

1. Subjective-Optimization: f_i is an optimal response to $\tilde{e}_i$;
2. Belief-Confirmation: $\mu_{f_i,e_i} = \mu_{f_i,\tilde{e}_i}$.

The beliefs a player holds at the beginning of the game, as described by his

subjective environment function $\tilde{e}_i$, may depend on past stochastic observations. This dependency was ignored in the single player model of Section 3 since it was assumed that the choice of $\tilde{e}_i$ already took this past experience into account. However, in the multi-person case, if past observations of different players are correlated, then it is useful to describe explicitly how they create correlation in the individual belief functions.

We do this, as before, by amending a correlation device (M,p) to the beginning of the game. The subjective correlated game will be described by the real correlated game (M,p,G^∞) but together with beliefs which are message dependent. Formally, for each player i and message m_i let $\tilde{e}_{m_i}$ describe a (subjective) environment response function of G^∞ conjectured by player i. We assume that each player chooses a strategies f_{m_i} as a best response to each $\tilde{e}_{m_i}$. These vectors of individual choices result in a vector of individual strategies $f = (f_1,...,f_n)$ in the game with correlation. We let e_{m_i} denote the real environment response function induced by the vector f on player i, conditional on his initial message m_i.

Definition 4.2: A subjective correlated equilibrium for G^∞ consists of a correlation device (M,p) as above, with a vector of strategies $f = (f_1,...,f_n)$ of (M,p,G^∞) and a vector of (subjective) environment response functions $\tilde{e} = (\tilde{e}_1,...,\tilde{e}_n)$ satisfying for each player i and message m_i the following two conditions:

1. Subjective-Optimization: f_{m_i} is a best response to $\tilde{e}_{m_i}$; and
2. Correlated Belief-Confirmation: $\mu_{f_{m_i},\tilde{e}_{m_i}} = \mu_{f_{m_i},e_{m_i}}$.

Clearly, every Nash equilibrium is a subjective Nash equilibrium, with $e_i = \tilde{e}_i$ for all i and, similarly, every correlated equilibrium is a subjective correlated equilibrium. However, the fact that the $\tilde{e}_i$'s may disagree with the e_i's off the play

path, makes the set of subjective equilibria significantly larger than the corresponding objective notions. Subjective Nash equilibria, which are not Nash equilibria, could be of economic interest of their own, as can be seen in the following example.

Example 4.2: Competitive Equilibrium is a Subjective Cournot Equilibrium With Finitely Many Producers. Consider a homogeneous-product repeated Cournot oligopoly game with n-identical producers. Each producer i has a constant marginal production cost of \$g/unit, with which he can produce any quantity a_i at any of the discrete times t = 1,2,... . The market price in each period is deterministic and linear, i.e., $p = b - d \Sigma_i a_i$ for some positive b and d with $b > g$.

Consider a vector of production levels $a^* = (a_1^*,...,a_n^*)$ resulting in a competitive market price p = g, i.e., $\Sigma a_i^* = (b - g)/d$. Suppose each player plays a constant strategy f_i^* which prescribes the constant production level a_i^* after every history. The vector of strategies $f^* = (f_1^*,...,f_n^*)$ is not a Nash equilibrium of the repeated game since each firm i is making a zero profit which could be increased by reducing production.

Nevertheless, the above production levels are supported by a subjective equilibrium of the repeated game, if each of the finitely many players assumes that he cannot affect the prices. For example, assume that the outcome reported to each player at the end of each period consists of his own production level and realized market price. Let each player hold beliefs described by the stationary subjective environment response function $\tilde{e}_i|_{h_i a_i}(g) = 1$. That is he assumes that with probability one the market price will be g regardless of past history of prices and regardless of his production level. Clearly, producing a_i^* is a best response to such $\tilde{e}_i$. Moreover, the price sequence (g,g,...) is assigned probability one by him and,

indeed, it has probability one under f. So f confirms the beliefs $\bar{e}$. Thus, we are in a subjective equilibrium.

It is easy to see in the above model that the only subjective equilibrium which is stationary in actions and beliefs is the competitive one. Thus, the only stationary subjective equilibrium in the Cournot game is the competitive one. This example illustrates that, while subjective equilibrium by itself may allow many outcomes in a game, in the presence of additional assumptions on beliefs it may lead to interesting conclusions.

Notice that, in the above discussion, the stationarity of beliefs could be significantly weakened provided that we keep each player believing that his actions do not alter the price distribution. An interesting case of this type is when each player believes that tomorrow's price will be what today's price was.

4.4 Convergence to Subjective Correlated Equilibrium

In the previous section we justified the notions of subjective Nash, and subjective correlated, equilibrium by arguing that players, finding themselves at such a situation, will have no reason to alter their beliefs or strategies. In this section we present sufficient conditions under which utility-maximizing players must eventually play a subjective correlated equilibrium.

Since the individual strategies, however, may not in general converge to a stationary limit strategy, we will follow the same course as we did in the one person case, Theorem 3.1. In other words, we will show that after sufficiently long finite time they must play a subjective correlated $\epsilon-$equilibrium for arbitrarily small ϵ.

Definition 4.4.1: Let (M,p,G^{∞}) be a correlated game, f a vector of correlated strategies, $\bar{e}$ a vector of correlated subjective environment functions, and $\epsilon > 0$. We

say that $(f,\tilde{e})$ is a subjective correlated $\epsilon-$equilibrium if the following conditions hold.

1. Subjective-Optimization: For every player i and message m_i, f_{m_i} is a best response to $\tilde{e}_{m_i}$.
2. Correlated $\epsilon-$Belief-Confirmation: With probability greater than $1-\epsilon$, a message vector m will be chosen with $\mu_{f_{m_i},\tilde{e}_{m_i}}$ being $\epsilon-$close to $\mu_{f_{m_i},e_{m_i}}$.

Before stating the convergence result we recall the terminology of Section 4.1. Let f be a vector of strategies of G^∞, e be the induced vector of environment response functions, $\tilde{e}$ be a vector of (subjective) environment response functions, and t a positive integer. The correlated game induced from time t is a correlated game (H^t,μ^t,G^∞), with H^t denoting all the possible histories of length t, and μ^t is μ_f restricted to the events in H^t. f^t, e^t and $\tilde{e}^t$ are the concepts induced on the correlated game by the original game in the natural way, as already discussed.

We say that the players play a subjective correlated $\epsilon-$equilibrium from time t on (correlated on the past) if $(f^t,\tilde{e}^t)$ is a subjective correlated $\epsilon-$equilibrium in the game (H^t,μ^t,G^∞).

Recalling the definition in Section 3, we say that $(f_i,\tilde{e}_i)$ is compatible with (f_i,e_i) if μ_{f_i,e_i} is absolutely continuous with respect to $\mu_{f_i,\tilde{e}_i}$. The following result is, mathematically, an immediate consequence of the convergence result for the one player case.

Theorem 4.4.1: Let f be a vector of strategies and $\tilde{e}$ be a vector of subjective environment functions. Suppose f and $\tilde{e}$ satisfy the following two conditions for

every player i:

1. Subjective Optimization: f_i is a best response to $\tilde{e}_i$, and
2. Beliefs Compatible with the Truth: $(f_i,\tilde{e}_i)$ is compatible with (f_i,e_i).

Then for every $\epsilon > 0$ there is a time T such that from all times t on, with $t \geq T$, the players play a subjective correlated ϵ-equilibrium.

Remark 4.4.1: Starting with a Correlated Game. It is easy to see that Theorem 4.4.1 can be extended to the case that the original strategies were correlated. That is, instead of playing G^{∞} directly, the player starts at time zero with the observation of some correlation device and choose their subjectively optimal strategies as best response to beliefs which are message dependent. The conclusion, that they will eventually play a subjective correlated equilibrium will be identical to the one in the statement of the current Theorem 4.4.1.

5. Coincidence of Subjective and Objective Equilibria

The convergence theorem of the previous section illustrates conditions that must lead the players to a subjective correlated equilibrium. The subjective notion of equilibrium is, in most cases, more plausible than the objective counterpart, but it entails a reduced prediction power. Since any player may hold his own individual hypothesis that justifies his actions, an outside analyst who wants to predict future outcomes must collect information about players' subjective beliefs. The potential contribution of subjective equilibrium to prediction power depends on the game and on players' beliefs. The preliminary examples given here illustrate situations, with general conditions on beliefs, involving no loss of prediction power when compared

to the objective equilibrium. That is, subjective and objective equilibrium predict the same behavior.

5.1. Optimistic and Pessimistic Conjectures

In the multi-arm bandit example discussed above, the player does not know the real distribution nature uses to determine his outcomes. A suboptimal arm may be employed whenever the payoff of other arms are underestimated. In other words, subjective equilibrium in this case is not an objective one, since pessimistic conjectures regarding unused arms are held.

The same logic extends to multi-player games, as demonstrated in Example 2.2. It is natural to expect that, if we rule out pessimistic beliefs, the behavior induced by a subjective equilibrium must coincide with the behavior induced by an objective one. For this purpose, and later ones, we need to introduce the following notions of equivalence of behavior.

Two strategy vectors f and g, of G^∞ (or a correlated version of it in (M,P,G^∞)), play like each other if the distributions they induce on the space play paths, Z^∞, coincide, i.e., $\mu_f = \mu_g$. Notice that when this is the case, using any statistical tools, none of the players could tell, after observing any initial segment of play, or even after watching the infinite play, whether f or g was played. Moreover, even an outside observer with the ability to perfectly monitor all players' actions could not distinguish between f and g. This is so because disagreements between f and g can only occur off the play path, thus, with probability zero.

A weaker concept, to be used later on, is when two strategies f and f′ are non-distinguishable to the players. This means, as in the notion of belief-confirmation, that the marginal distributions, on the individual play paths of each player, coincide. Formally, let e_i and e_i' be player i's environment response functions

induced by f_{-i} and f'_{-i}, respectively, we require that $\mu_{f_i,e_i} = \mu_{f'_i,e'_i}$. If f and f' are non-distinguishable to the players, they induce the same marginal distributions on each player's payoff paths, and thus yield the same utility to each player. However, an outsider with full monitoring power could observe events to which f and f' assign different probability. Our present results use the stronger concept but the weaker one can be used in other examples.

Let f be a vector of strategies of the infinite game, with or without correlation, and let e_i be the induced environment response function. We say that $\tilde{e}_i$ has optimistic conjectures relative to f if for every strategy g_i, $u_i(g_i,e_i) \leq u_i(g_i,\tilde{e}_i)$.

The following proposition states that whenever agents are optimistic, any subjective equilibrium is an objective equilibrium.

Proposition 5.1: Let $(f,\tilde{e})$ be a subjective correlated (resp. Nash) equilibrium with each $\tilde{e}_i$ holding optimistic conjectures relative to f. Then f is a correlated (resp. Nash) equilibrium.

Proof: Suppose to the contrary that f_i is not optimal against the real e_i. Therefore, there exists a strategy g_i of player i satisfying $u_i(g_i,e_i) > u_i(f_i,e_i)$. However, by the optimistic conjecture assumption, $u_i(g_i,\tilde{e}_i) \geq u(g_i,e_i)$. Moreover, since $(f,\tilde{e})$ is a subjective equilibrium, $u_i(f_i,e_i) = u_i(f_i,\tilde{e}_i)$. As we combine the first two inequalities with the last equality we get $u_i(g_i,\tilde{e}_i) > u_i(f_i,\tilde{e}_i)$, which contradicts the optimality of f_i against $\tilde{e}_i$. This concludes the proof. //

That each player holds subjective beliefs, an $\tilde{e}_i$, with optimistic conjectures relative to the actual play is a strong requirement. Yet without imposing some conditions that make players' conjectures realistic--or, better, yet, optimistic--one

may sustain any behavior by a subjective equilibrium. However, the following familiar economic model illustrates that, under some general assumptions, subjective and objective equilibria generate the same behavior pattern.

Example 5.2. Subjective Cournot Equilibrium Plays Like Cournot Equilibrium: We consider n-producers (players) of an identical product in a market with a commonly known downward sloping demand function, $D(\bullet)$.

To fit our model, and to simplify the exposition, we make the following assumptions. The set of outcomes, prices in this case, consists of all nonnegative rationals. Thus, C_i is a countable set for $i = 1,...,n$. Similarly, we let all players have the same set of actions, feasible production levels, $A_i = \{0,1,2,...\}$. We assume that in each period the market price is established deterministically according to the vector of production levels, $a = (a_1,...,a_n)$, by $c = D(\Sigma\, a_i)$. We also assume for simplicity that they each have a constant and positive marginal production costs, K. So if in a given period a player produces at a level a_i and the realized market price (determined by all production levels) is c, his period net profit is a_ic,a_iK.

We let (M,P,G^∞) be the above game with some initial correlation device (M,P) describing the distribution of information available to the players prior to the start of the game. We assume that $(f,\tilde{e})$ is a subjective correlated equilibrium. Thus, each f_{m_i} is a best response to $\tilde{e}_{m_i}$ and $\mu_{f_{m_i},\tilde{e}_{m_i}} = \mu_{f_{m_i},e_{m_i}}$, where e_{m_i} is the real environment response function determined by the game and the other player's strategies, given m_i, while $\tilde{e}_{m_i}$ is the one induced by the subjective conjecture of i.

We assume that each player knows the demand function. Formally, we do it by assuming that for every Δ the distribution $\tilde{e}_{m_i}|_{h_i(a_i+\Delta)}$ coincides with the distribution $D\left(D^{-1}\left(\tilde{e}_{m_i}|_{h_ia_i}\right) + \Delta\right)$. Notice that this rules out the price taking assumption we used to obtain the competitive prices at a subjective equilibrium.

Our goal now is to show that f plays like some g which is a correlated equilibrium, or equivalently, a Nash equilibrium of (M,P,G^{∞}).

We construct $g = (g_1,...,g_n)$ as follows. For each player i, after every history, h_i, which has a μ_f−positive probability we define g_i to coincide with f_i, i.e., $g_i|_{h_i} = f_i|_{h_i}$. Notice that this implies that g plays like f. For μ_f zero probability histories, h_i, define $g_i|_{h_i}$ to choose a large production level L with probability one. The level L is chosen in such a way that the amount $(n - 1)L$, produced by $n - 1$ producers, lowers market price below the marginal cost, K.

By the definition of g for every player i, f and g play alike. That is, player i cannot tell the difference between f and g because they induce the same distribution over the signals observed by player i.

In order to show that g is an equilibrium, we show that for every possible deviation, g_i', of player i, $u_i(g_i',e_i) \leq u_i(g_i,\tilde{e}_i)$, where e_i is the environment function induced by g_{-i}.

We will show that g_i' is not a profitable deviation by showing that the outcome generated by (g_i',e_i) could be generated by some f_i' and $\tilde{e}_i$. Since f_i is individually optimal (against $\tilde{e}_i$), $u_i(f_i,\tilde{e}_i) \geq u_i(f_i',\tilde{e}_i) = u_i(g_i',e_i)$. As $u_i(f_i,\tilde{e}_i) = u_i(g_i,e_i)$ because f and g play alike, we conclude that $u_i(g,e_i) \geq u_i(g_i',e_i)$.

Recall the demand function, D, is commonly known and that it has a negative slope. Suppose that after the history h_i the strategy f_i prescribes player i the action a_i with positive probability. Since f is a subjective equilibrium player i knows to predict the distribution over prices given his own a_i. As D is one-to-one, player i is able to forecast after h_i the distribution over the quantity produced by all his competitors. Therefore, he knows to predict the distribution over prices not only given a_i, but also given any other quantity player i may produce. We now deduce that after every history with positive probability (w.r.t. to f) player i knows the

distribution over prices induced by (g_i',e_i). In other words, the distribution over prices induced by (g_i',e_i) is the one induced by $(g_i',\tilde{e}_i)$ after every history with positive probability. By iterating the same argument we infer that the probability assigned to history h_i by $(g_i',\tilde{e}_i)$ and by (g_i',e_i) is the same, provided that h_i has a $(g_i,\tilde{e}_i)$-positive probability.

Define $f_i'(h_i)$ to be identical to $g_i(h_i)$ for every history h_i which is positive w.r.t. $(g_i,\tilde{e}_i)$. Otherwise, $f_i'(h_i)$ is zero. We will show that $u_i(f_i',\tilde{e}_i) \geq u_i(g_i',e_i)$. Fix a time t and let h_i be a history of length t. Conditioned on h_i being $(g_i,\tilde{e}_i)$-positive we get that $u_i(f_i',\tilde{e}_i)$, which by definition equals $u_i(g_i',\tilde{e}_i)$, is equal to $u_i(g_i',e_i)$. As for every other history h_i, since $f_i(h_i) = 0$, the return for player i is zero. On the other hand, the payoff $u_i(g_i,e_i)$ is at most zero (after h_i) because the total amount produced by all players drops market price below the cost per unit. Therefore, conditioned on h_i being a history with probability zero (w.r.t $(g_i,\tilde{e}_i)$) $0 = u_i(f_i',\tilde{e}_i) \geq u_i(g_i,e_i)$.

We may conclude that at any period t $u_i(f_i',\tilde{e}_i) \geq u_i(g_i,e_i)$ and therefore this is the case for the whole repeated game. This completes the proof, showing that g is an equilibrium.

6. Subjective Extensive Form Games

In the previous sections we restricted the subjective approach to repeated stochastic-outcome games. The extension to general extensive form games is straightforward.

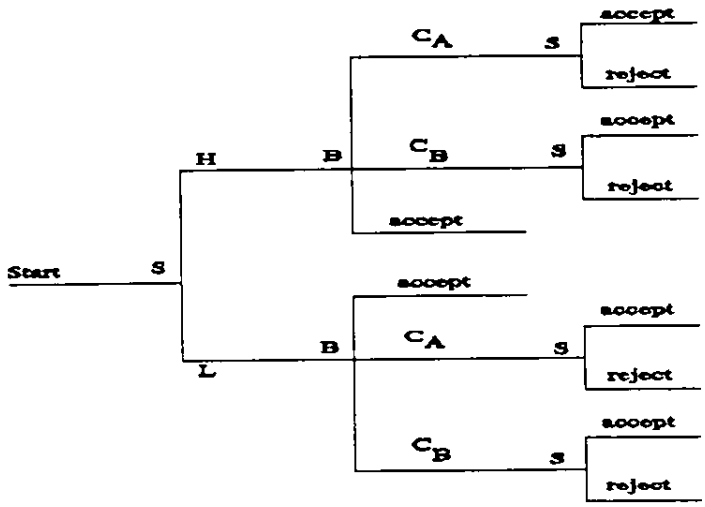
We need only to modify the definition of the environment response functions. Recall that $e_{i|h_i a_i}(c_i)$ represented the probability that outcome c_i will be realized by player i after the play consisting of the individual history h_i followed by the action a_i. The role of h_i's in the above must be replaced by the player's information sets. For every information set h_i, the a_i's following it must be restricted to actions feasible at

this information set. The consequences, c_i's in the above, can be replaced by two objects. They could be terminal nodes with their associated payoffs, when the action a_i taken at h_i can lead to such termination. They could also describe new individual information sets if following $h_i a_i$ the other players could lead player i to a "next" information set.

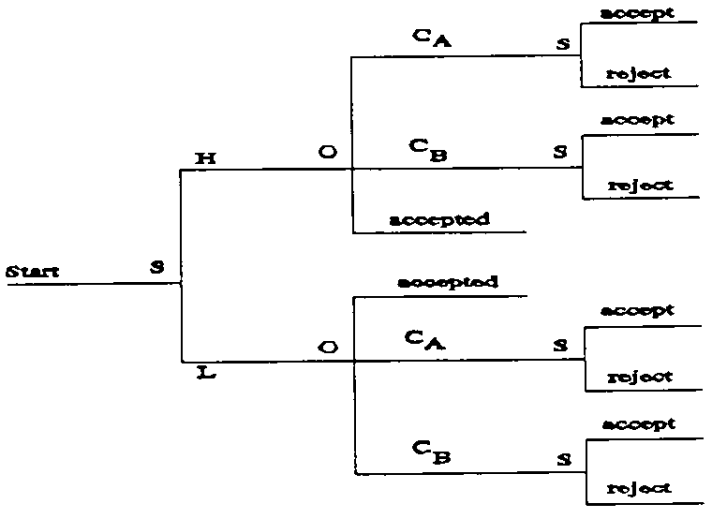
The following example illustrates such a generalization and its usefulness without the need to develop the general terminology of extensive form games.

Consider a three stage alternating offer bargaining between a seller and a buyer. At stage one the seller can ask the buyer for two prices, H or L. In the second stage, the buyer can accept the asked price, X, with X = H or X = L, yielding the respective payoffs $X - R_s$, $R_b - X$, where R_s and R_b represent the respective reservation values. Or the buyer can counter propose two prices, C_A or C_B, which the seller than accepts or rejects. If C_X is accepted, X = A or X = B. Again, the respective payoffs are $C_X - R_s$, $R_b - C_X$. If it is declined, the resulting payoffs are 0 and 0.

The extensive form game has the following simple representation.



The decision tree of the seller has the following structure.



His subjective environment response function will specify six probabilities corresponding to the six arcs marked accepted, C_A and C_B. For example, $e_s|_L$ (accepted) represents the probability that a seller's initial low offer of L will be accepted, and $e_s|_H(C_A)$ represents the probability that an initial H offer will be

responded to with a counteroffer C_A.

However, different games will give rise to the same decision tree of the seller as illustrated by the following two scenarios.

Scenario 1: The buyer consists of two players, b_1 and b_2, with hierarchical decision making. Upon hearing the asked price X, player b_1 can accept, counter-propose C_A, counter-propose C_B, or pass the decision to b_2. If b_1 passes, then b_2 decides whether to accept, or counter with C_A or C_B. Now there are three reservation values, and if the item is sold at price P, the respective payoffs are ($P - R_S$, $R_{b_1} - P$, $R_{b_2} - P$).

Notice that the only concern to the seller are the probabilities given by his environment response function and not the process of the decision making by the group of buyers. We could construct a large number of scenarios, like the one above, all of which constitute different game trees with different possible interacting buyers, but all yielding the same individual decision tree for the seller.

Scenario 2: A Bayesian game with unknown buyer's reservation value. Suppose the single buyer has two possible reservation values chosen according to some prior probabilities. The buyer knows the realized value, but the seller does not. Now we have two versions of the original game, with nature moving first and choosing which of the two trees to enter. The buyer knows which tree natures chooses, but the seller does not. Every pair of corresponding nodes in the two trees are put together for him in a single information set.

Nevertheless, his individual decision tree is unchanged and all he cares to know are the probabilities of his offers being accepted or countered.

When we combine variations, as in Scenarios 1 and 2 together, we see that there is a large number of games, all yielding the seller the same individual decision tree, and all he needs are the assessments in this one decision tree of the probability of various responses to his offers.

The buyer has a similar task. The real game induces on him a known decision tree. To decide on his optimal strategy, he needs to assess probabilities of how the seller (or sellers or sellers-agents) will respond to his counter offers.

The subjective extensive game consists of the real game together with the two decision trees and individually assessed environment response functions.

A subjectively optimal strategy consists of optimal choices in the individual decision tree relative to the assessed responses. A vector of such strategies is a subjective equilibrium if all subjective forecasts over future individual outcomes are accurate. Assessments off the play path do not have to be accurate.

For example, suppose the seller assesses probabilities 1 to his proposed L price being countered with C_A. He also holds assessments regarding the probabilities of response to his proposal H price which leads him to a subjective optimal strategy of proposing L and accepting any counter proposal. The buyer, due to his own assessments, chooses an optimal strategy which indeed counters L with a C_A offer. If the buyer's beliefs are that the seller will offer L and accept a counter proposal C_A, then we are at a subjective equilibrium. Every player's expectations on the play path are met, even though their conjectures regarding off path responses may be wrong.

The above subjective equilibrium is closely related to earlier examples in Fudenberg and Kreps (1988) and to self-confirming equilibrium in the Fudenberg-Levine (1993) sense.

A major difference is that the players are not assumed to know the game.

For example, the real game could be as in Scenario 1 above, yet with the seller behaving and if he is in the original game facing a single buyer.

Another interesting discrepancy between the players' model of the game and the actual game may be regarding the continuation of the game. The buyer may think that the game may continue with additional offers and counter-offers, yet the seller thinking that he must make a final response to the buyer's counter-offer. Since at the equilibrium proposal above the game ends after the buyer's counter-offer, whichever of them is wrong regarding the possibility or impossibility of continuation will never find out.

References

Abreu, D., D. Pearce and E. Staccheti (1986), "Toward A Theory of Discounted Repeated Games with Imperfect Monitoring," Econometrica, 58, 1041-1064.

Aumann, R. J. (1974), "Subjectivity and Correlation in Randomized Strategies," Journal of Mathematical Economics, 67-96.

Aumann, R. J. (1987), Correlated Equilibrium as an Expression of Bounded Rationality," Econometrica, 55, 1-19.

Banks, J. S. and R. K. Sundaran (1993), "Switching Costs and the Giffins Index," Working Paper No. 353, University of Rochester.

Battigalli, P. (1987), "Comportamento Razionale ed Equilibrio nei Giochi e nelle Situazioni Sociali," Unpublished Dissertation, Bocconi University, Milano.

Battigalli, P. and D. Guaitoli (1988), "Conjectured Equilibria and Rationalizability in a Macroeconomic Game with Incomplete Information," Bocconi University.

Battigalli, P., M. Gilli, and M. C. Molinari (1992), "Learning Convergence to Equilibrium in Repeated Strategic Interactions: An Introductory Survey," Ricerche Economiche, forthcoming.

Bernheim, D. (1986), "Axiomatic Characterizations of Rational Choice in Strategic Environments," Scandanavian Journal of Economics, 88, 473-488.

Blackwell, D. and L. Dubins (1962), "Merging of Opinions with Increasing Information," Annals of Mathematical Statistics, 38, 882-886.

Blume, L. and D. Easley (1992), "Rational Expectations and Rational Learning," Cornell University.

Chichilnsky, G. (1992), "Existence and Optimality of a General Equilibrium with Endogenous Uncertainty," Columbia University, mimeo.

Crawford, V. and H. Haller (1990), "Learning How to Cooperate: Optimal Play in

Repeated Coordination Games," Econometrica, 58, 571-596.

Cripps, M. and J. Thomas (1991), "Learning and Reputation in Repeated Games of Incomplete Information," University of Warwick.

El-Gamal, M. (1992), "The Rational Expectations of ϵ-Equilibrium," California Institute of Technology Social Science Working Paper No. 823.

Fudenberg, D. and D. Kreps (1988), "A Theory of Learning, Experimentation and Equilibrium in Games," Stanford University.

Fudenberg, D. and D. Levine (1993), "Self-Confirming Equilibrium," Econometrica, 61, 523-545.

Fudenberg, D. and J. Tirole (1992), Game Theory, MIT Press.

Fujiwara-Grew, T. (1993), "A Note on Kalai-Lehrer Learning Model with a Generalized Semi-Standard Information," Stanford University.

Gilboa, I. and D. Schmeidler (1992), "Case-Based Decision Theory," Northwestern University.

Goyal, S. and M. Janssen (1993), "Can We Rationally Learn to Coordinate?" Department of Economics, Erasmus University, Rotterdam.

Green, E. J. and R. H. Porter (1984), "Noncooperative Collusion Under Imperfect Price Information," Econometrica, 52, 87-100.

Hahn, F. (1973), "On the Notion of Equilibrium in Economics: An Inaugural Lecture," Cambridge: Cambridge University Press.

Harsanyi, J. C. (1967), "Games of Incomplete Information Played by Bayesian Players, Part I," Management Science, 14, 159-182.

Hayek, F. A., von (1937), "Economics of Knowledge," Economica, 4, 33-54.

Jordan, J. S. (1991), "Bayesian Learning in Normal Form Games," Games and Economic Behavior, 3, 60-81.

Jordan, J. S. (1992), "The Exponential Convergence of Bayesian Learning in Normal

Form Games," Games and Economic Behavior, 4, 202-217.

Jordan, J. (1993), "Three Problems in Learning Mixed-Strategy Nash Equilibria," Games and Economic Behavior, 5 (3), 368-386.

Kalai, E. and E. Lehrer (1990a), "Bayesian Learning and Nash Equilibrium," Northwestern University.

Kalai, E. and E. Lehrer (1993a), "Subjective Equilibrium in Repeated Games," Econometrica, 61, 1231-1240.

Kalai, E. and E. Lehrer (1993b), "Rational Learning Leads to Nash Equilibrium," Econometrica, 61, 1019-1045.

Kalai, E. and E. Lehrer (1990c), "Weak and Strong Merging of Opinions," Journal of Mathematical Economics, forthcoming.

Koutsougeras, L. C. and N. C. Yannelis (1993), "Convergence and Approximate Results for Non-Cooperative Bayesian Games: Learning Theorems," Economic Theory, forthcoming.

Kuhn, H. W. (1953), "Extensive Games and the Problem of Information," in H. W. Kuhn and A. W. Tucker (eds.), Contributions to the Theory of Games, Vol. II, pp. 193-216, Annals of Mathematics Studies, 28, Princeton University Press.

Kurz, M. (1994), "General Equilibrium with Endogenous Uncertainty," to appear in On the Formulation of Economic Theory, G. Chichilnisky (ed.), Cambridge University Press.

Lehrer, E. (1991), "Internal Correlation in Repeated Games," International Journal of Game Theory, 19, 431-456.

Lehrer, E. and R. Smorodinsky (1993), "Compatible Measures and Merging," mimeo.

Mertens, J. F. and S. Zamir (1985), "Formalization of Bayesian Analysis for Games Incomplete Information," International Journal of Game Theory, 14, 1-29.

Monderer, D. and D. Samet (1990), "Stochastic Common Learning," Games and Economic Behavior, forthcoming.

Myerson, R. (1991), Game Theory: Analysis of Conflict, Harvard University Press.

Nash, J. F. (1950), "Equilibrium Points in n-person Games," Proceedings of the National Academy of Sciences USA, 36, pp. 48-49.

Nyarko, Y. (1991a), "The Convergence of Bayesian Belif Hierarchies," C. V. Starr Center Working Paper No. 91-50, New York University.

Nyarko, Y. (1991b), "Bayesian Learning Without Common Priors and Convergence to Nash Equilibrium," New York University.

Pearce D. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection," Econometrica, 52, 1029-1050.

Porter, R. H. (1983), "Optimal Cartel Trigger-Price Strategies," Journal of Economic Theory, 29, 313-338.

Rothschild, M. (1974), "A Two-Armed Bandit Theory of Market Pricing," Journal of Economic Theory, 9, 195-202.

Rubinstein, A. and A. Wolinsky (1990), "Rationalizable Conjectural Equilibrium: Between Nash and Rationalizability," Games and Economic Behavior, forthcoming.

Schmidt, K. M. (1991), "Reputation and Equilibrium Characterization in Repeated Games of Conflicting Interest," Bonn University.

Selten, R. (1975), "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory, 4, 25-55.

Vives, X. (1992), "How Fast Do Rational Agents Learn?" Review of Economic Studies, forthcoming.

Watson, J. (1992), "Reputation and Outcome Selection in Perturbed Supergames: An Intuitive, Behavioral Approach," Stanford University.

Wittle, P. (1982), Optimization Over Time, Vol. 1, Wiley, New York.